To display an HTML page correctly, a web browser must know the character set (character encoding) to use.

**What is Character Encoding?**

ASCII was the first **character encoding standard** (also called character set). It defines 127 different alphanumeric characters that could be used on the internet.

ASCII supported numbers (0-9), English letters (A-Z), and some special characters like ! $ + - ( ) @ < > .

ANSI (Windows-1252) was the original Windows character set. It supported 256 different character codes.

ISO-8859-1 was the default character set for HTML 4. It also supported 256 different character codes.

Because ANSI and ISO were limited, the default character encoding was changed to UTF-8 in HTML5.

UTF-8 (Unicode) covers almost all of the characters and symbols in the world.

**\*\*\*Note:** All HTML 4 processors also support UTF-8.

**The HTML charset Attribute**

To display an HTML page correctly, a web browser must know the character set used in the page.

This is specified in the <meta> tag:

**For HTML4:**

<meta http-equiv="Content-Type" content="text/html;charset=ISO-8859-1">

**For HTML5:**

<meta charset="UTF-8">

**\*\*\*Note:** If a browser detects ISO-8859-1 in a web page, it defaults to ANSI, because ANSI is identical to ISO-8859-1 except that ANSI has 32 extra characters.

**Differences Between Character Sets**

The following table displays the differences between the character sets described above:

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Numb** | **ASCII** | **ANSI** | **8859** | **UTF-8** | **Description** |
| 32 |  |  |  |  | space |
| 33 | ! | ! | ! | ! | exclamation mark |
| 34 | " | " | " | " | quotation mark |
| 35 | # | # | # | # | number sign |
| 36 | $ | $ | $ | $ | dollar sign |
| 37 | % | % | % | % | percent sign |
| 38 | & | & | & | & | ampersand |
| 39 | ' | ' | ' | ' | apostrophe |
| 40 | ( | ( | ( | ( | left parenthesis |
| 41 | ) | ) | ) | ) | right parenthesis |
| 42 | \* | \* | \* | \* | asterisk |
| 43 | + | + | + | + | plus sign |
| 44 | , | , | , | , | comma |
| 45 | - | - | - | - | hyphen-minus |
| 46 | . | . | . | . | full stop |
| 47 | / | / | / | / | solidus |
| 48 | 0 | 0 | 0 | 0 | digit zero |
| 49 | 1 | 1 | 1 | 1 | digit one |
| 50 | 2 | 2 | 2 | 2 | digit two |
| 51 | 3 | 3 | 3 | 3 | digit three |
| 52 | 4 | 4 | 4 | 4 | digit four |
| 53 | 5 | 5 | 5 | 5 | digit five |
| 54 | 6 | 6 | 6 | 6 | digit six |
| 55 | 7 | 7 | 7 | 7 | digit seven |
| 56 | 8 | 8 | 8 | 8 | digit eight |
| 57 | 9 | 9 | 9 | 9 | digit nine |
| 58 | : | : | : | : | colon |
| 59 | ; | ; | ; | ; | semicolon |
| 60 | < | < | < | < | less-than sign |
| 61 | = | = | = | = | equals sign |
| 62 | > | > | > | > | greater-than sign |
| 63 | ? | ? | ? | ? | question mark |
| 64 | @ | @ | @ | @ | commercial at |
| 65 | A | A | A | A | Latin capital letter A |
| 66 | B | B | B | B | Latin capital letter B |
| 67 | C | C | C | C | Latin capital letter C |
| 68 | D | D | D | D | Latin capital letter D |
| 69 | E | E | E | E | Latin capital letter E |
| 70 | F | F | F | F | Latin capital letter F |
| 71 | G | G | G | G | Latin capital letter G |
| 72 | H | H | H | H | Latin capital letter H |
| 73 | I | I | I | I | Latin capital letter I |
| 74 | J | J | J | J | Latin capital letter J |
| 75 | K | K | K | K | Latin capital letter K |
| 76 | L | L | L | L | Latin capital letter L |
| 77 | M | M | M | M | Latin capital letter M |
| 78 | N | N | N | N | Latin capital letter N |
| 79 | O | O | O | O | Latin capital letter O |
| 80 | P | P | P | P | Latin capital letter P |
| 81 | Q | Q | Q | Q | Latin capital letter Q |
| 82 | R | R | R | R | Latin capital letter R |
| 83 | S | S | S | S | Latin capital letter S |
| 84 | T | T | T | T | Latin capital letter T |
| 85 | U | U | U | U | Latin capital letter U |
| 86 | V | V | V | V | Latin capital letter V |
| 87 | W | W | W | W | Latin capital letter W |
| 88 | X | X | X | X | Latin capital letter X |
| 89 | Y | Y | Y | Y | Latin capital letter Y |
| 90 | Z | Z | Z | Z | Latin capital letter Z |
| 91 | [ | [ | [ | [ | left square bracket |
| 92 | \ | \ | \ | \ | reverse solidus |
| 93 | ] | ] | ] | ] | right square bracket |
| 94 | ^ | ^ | ^ | ^ | circumflex accent |
| 95 | \_ | \_ | \_ | \_ | low line |
| 96 | ` | ` | ` | ` | grave accent |
| 97 | a | a | a | a | Latin small letter a |
| 98 | b | b | b | b | Latin small letter b |
| 99 | c | c | c | c | Latin small letter c |
| 100 | d | d | d | d | Latin small letter d |
| 101 | e | e | e | e | Latin small letter e |
| 102 | f | f | f | f | Latin small letter f |
| 103 | g | g | g | g | Latin small letter g |
| 104 | h | h | h | h | Latin small letter h |
| 105 | i | i | i | i | Latin small letter i |
| 106 | j | j | j | j | Latin small letter j |
| 107 | k | k | k | k | Latin small letter k |
| 108 | l | l | l | l | Latin small letter l |
| 109 | m | m | m | m | Latin small letter m |
| 110 | n | n | n | n | Latin small letter n |
| 111 | o | o | o | o | Latin small letter o |
| 112 | p | p | p | p | Latin small letter p |
| 113 | q | q | q | q | Latin small letter q |
| 114 | r | r | r | r | Latin small letter r |
| 115 | s | s | s | s | Latin small letter s |
| 116 | t | t | t | t | Latin small letter t |
| 117 | u | u | u | u | Latin small letter u |
| 118 | v | v | v | v | Latin small letter v |
| 119 | w | w | w | w | Latin small letter w |
| 120 | x | x | x | x | Latin small letter x |
| 121 | y | y | y | y | Latin small letter y |
| 122 | z | z | z | z | Latin small letter z |
| 123 | { | { | { | { | left curly bracket |
| 124 | | | | | | | | | vertical line |
| 125 | } | } | } | } | right curly bracket |
| 126 | ~ | ~ | ~ | ~ | tilde |
| 127 | DEL |  |  |  |  |
| 128 |  | € |  |  | euro sign |
| 129 |  |  |  |  | NOT USED |
| 130 |  | ‚ |  |  | single low-9 quotation mark |
| 131 |  | ƒ |  |  | Latin small letter f with hook |
| 132 |  | „ |  |  | double low-9 quotation mark |
| 133 |  | … |  |  | horizontal ellipsis |
| 134 |  | † |  |  | dagger |
| 135 |  | ‡ |  |  | double dagger |
| 136 |  | ˆ |  |  | modifier letter circumflex accent |
| 137 |  | ‰ |  |  | per mille sign |
| 138 |  | Š |  |  | Latin capital letter S with caron |
| 139 |  | ‹ |  |  | single left-pointing angle quotation mark |
| 140 |  | Œ |  |  | Latin capital ligature OE |
| 141 |  |  |  |  | NOT USED |
| 142 |  | Ž |  |  | Latin capital letter Z with caron |
| 143 |  |  |  |  | NOT USED |
| 144 |  |  |  |  | NOT USED |
| 145 |  | ‘ |  |  | left single quotation mark |
| 146 |  | ’ |  |  | right single quotation mark |
| 147 |  | “ |  |  | left double quotation mark |
| 148 |  | ” |  |  | right double quotation mark |
| 149 |  | • |  |  | bullet |
| 150 |  | – |  |  | en dash |
| 151 |  | — |  |  | em dash |
| 152 |  | ˜ |  |  | small tilde |
| 153 |  | ™ |  |  | trade mark sign |
| 154 |  | š |  |  | Latin small letter s with caron |
| 155 |  | › |  |  | single right-pointing angle quotation mark |
| 156 |  | œ |  |  | Latin small ligature oe |
| 157 |  |  |  |  | NOT USED |
| 158 |  | ž |  |  | Latin small letter z with caron |
| 159 |  | Ÿ |  |  | Latin capital letter Y with diaeresis |
| 160 |  |  |  |  | no-break space |
| 161 |  | ¡ | ¡ | ¡ | inverted exclamation mark |
| 162 |  | ¢ | ¢ | ¢ | cent sign |
| 163 |  | £ | £ | £ | pound sign |
| 164 |  | ¤ | ¤ | ¤ | currency sign |
| 165 |  | ¥ | ¥ | ¥ | yen sign |
| 166 |  | ¦ | ¦ | ¦ | broken bar |
| 167 |  | § | § | § | section sign |
| 168 |  | ¨ | ¨ | ¨ | diaeresis |
| 169 |  | © | © | © | copyright sign |
| 170 |  | ª | ª | ª | feminine ordinal indicator |
| 171 |  | « | « | « | left-pointing double angle quotation mark |
| 172 |  | ¬ | ¬ | ¬ | not sign |
| 173 |  |  |  |  | soft hyphen |
| 174 |  | ® | ® | ® | registered sign |
| 175 |  | ¯ | ¯ | ¯ | macron |
| 176 |  | ° | ° | ° | degree sign |
| 177 |  | ± | ± | ± | plus-minus sign |
| 178 |  | ² | ² | ² | superscript two |
| 179 |  | ³ | ³ | ³ | superscript three |
| 180 |  | ´ | ´ | ´ | acute accent |
| 181 |  | µ | µ | µ | micro sign |
| 182 |  | ¶ | ¶ | ¶ | pilcrow sign |
| 183 |  | · | · | · | middle dot |
| 184 |  | ¸ | ¸ | ¸ | cedilla |
| 185 |  | ¹ | ¹ | ¹ | superscript one |
| 186 |  | º | º | º | masculine ordinal indicator |
| 187 |  | » | » | » | right-pointing double angle quotation mark |
| 188 |  | ¼ | ¼ | ¼ | vulgar fraction one quarter |
| 189 |  | ½ | ½ | ½ | vulgar fraction one half |
| 190 |  | ¾ | ¾ | ¾ | vulgar fraction three quarters |
| 191 |  | ¿ | ¿ | ¿ | inverted question mark |
| 192 |  | À | À | À | Latin capital letter A with grave |
| 193 |  | Á | Á | Á | Latin capital letter A with acute |
| 194 |  | Â | Â | Â | Latin capital letter A with circumflex |
| 195 |  | Ã | Ã | Ã | Latin capital letter A with tilde |
| 196 |  | Ä | Ä | Ä | Latin capital letter A with diaeresis |
| 197 |  | Å | Å | Å | Latin capital letter A with ring above |
| 198 |  | Æ | Æ | Æ | Latin capital letter AE |
| 199 |  | Ç | Ç | Ç | Latin capital letter C with cedilla |
| 200 |  | È | È | È | Latin capital letter E with grave |
| 201 |  | É | É | É | Latin capital letter E with acute |
| 202 |  | Ê | Ê | Ê | Latin capital letter E with circumflex |
| 203 |  | Ë | Ë | Ë | Latin capital letter E with diaeresis |
| 204 |  | Ì | Ì | Ì | Latin capital letter I with grave |
| 205 |  | Í | Í | Í | Latin capital letter I with acute |
| 206 |  | Î | Î | Î | Latin capital letter I with circumflex |
| 207 |  | Ï | Ï | Ï | Latin capital letter I with diaeresis |
| 208 |  | Ð | Ð | Ð | Latin capital letter Eth |
| 209 |  | Ñ | Ñ | Ñ | Latin capital letter N with tilde |
| 210 |  | Ò | Ò | Ò | Latin capital letter O with grave |
| 211 |  | Ó | Ó | Ó | Latin capital letter O with acute |
| 212 |  | Ô | Ô | Ô | Latin capital letter O with circumflex |
| 213 |  | Õ | Õ | Õ | Latin capital letter O with tilde |
| 214 |  | Ö | Ö | Ö | Latin capital letter O with diaeresis |
| 215 |  | × | × | × | multiplication sign |
| 216 |  | Ø | Ø | Ø | Latin capital letter O with stroke |
| 217 |  | Ù | Ù | Ù | Latin capital letter U with grave |
| 218 |  | Ú | Ú | Ú | Latin capital letter U with acute |
| 219 |  | Û | Û | Û | Latin capital letter U with circumflex |
| 220 |  | Ü | Ü | Ü | Latin capital letter U with diaeresis |
| 221 |  | Ý | Ý | Ý | Latin capital letter Y with acute |
| 222 |  | Þ | Þ | Þ | Latin capital letter Thorn |
| 223 |  | ß | ß | ß | Latin small letter sharp s |
| 224 |  | à | à | à | Latin small letter a with grave |
| 225 |  | á | á | á | Latin small letter a with acute |
| 226 |  | â | â | â | Latin small letter a with circumflex |
| 227 |  | ã | ã | ã | Latin small letter a with tilde |
| 228 |  | ä | ä | ä | Latin small letter a with diaeresis |
| 229 |  | å | å | å | Latin small letter a with ring above |
| 230 |  | æ | æ | æ | Latin small letter ae |
| 231 |  | ç | ç | ç | Latin small letter c with cedilla |
| 232 |  | è | è | è | Latin small letter e with grave |
| 233 |  | é | é | é | Latin small letter e with acute |
| 234 |  | ê | ê | ê | Latin small letter e with circumflex |
| 235 |  | ë | ë | ë | Latin small letter e with diaeresis |
| 236 |  | ì | ì | ì | Latin small letter i with grave |
| 237 |  | í | í | í | Latin small letter i with acute |
| 238 |  | î | î | î | Latin small letter i with circumflex |
| 239 |  | ï | ï | ï | Latin small letter i with diaeresis |
| 240 |  | ð | ð | ð | Latin small letter eth |
| 241 |  | ñ | ñ | ñ | Latin small letter n with tilde |
| 242 |  | ò | ò | ò | Latin small letter o with grave |
| 243 |  | ó | ó | ó | Latin small letter o with acute |
| 244 |  | ô | ô | ô | Latin small letter o with circumflex |
| 245 |  | õ | õ | õ | Latin small letter o with tilde |
| 246 |  | ö | ö | ö | Latin small letter o with diaeresis |
| 247 |  | ÷ | ÷ | ÷ | division sign |
| 248 |  | ø | ø | ø | Latin small letter o with stroke |
| 249 |  | ù | ù | ù | Latin small letter u with grave |
| 250 |  | ú | ú | ú | Latin small letter u with acute |
| 251 |  | û | û | û | Latin small letter with circumflex |
| 252 |  | ü | ü | ü | Latin small letter u with diaeresis |
| 253 |  | ý | ý | ý | Latin small letter y with acute |
| 254 |  | þ | þ | þ | Latin small letter thorn |
| 255 |  | ÿ | ÿ | ÿ | Latin small letter y with diaeresis |

**The ASCII Character Set**

ASCII uses the values from 0 to 31 (and 127) for control characters.

ASCII uses the values from 32 to 126 for letters, digits, and symbols.

ASCII does not use the values from 128 to 255.

**The ANSI Character Set (Windows-1252)**

ANSI is identical to ASCII for the values from 0 to 127.

ANSI has a proprietary set of characters for the values from 128 to 159.

ANSI is identical to UTF-8 for the values from 160 to 255.

**The ISO-8859-1 Character Set**

8859-1 is identical to ASCII for the values from 0 to 127.

8859-1 does not use the values from 128 to 159.

8859-1 is identical to UTF-8 for the values from 160 to 255.

**The UTF-8 Character Set**

UTF-8 is identical to ASCII for the values from 0 to 127.

UTF-8 does not use the values from 128 to 159.

UTF-8 is identical to both ANSI and 8859-1 for the values from 160 to 255.

UTF-8 continues from the value 256 with more than 10 000 different characters.